

# 과학기술보고서의 상호교환을 위한

## 메타데이터

RFC1807은 네트워크상에서 기술보고서를 원활하게 교환하기 위한 서지레코드에 대한 형식을 규정하고 있다. 이 형식은 MARC처럼 복잡하지 않고 자동으로 파스(parse)하기 쉽도록 설계되어 코넬대학의 Dienst 프로토콜과 스탠포드대학의 SIFT시스템 등에서 활발하게 사용되고 있는 메타데이터이다.

안계성/ 한국데이터베이스진흥센터 정책연구과 과장

### 1. 배경

RFC(Request For Comments)는 인터넷과 관련된 문서로서, 특히 네트워크의 프로토콜, 절차, 프로그램, 개념 등에 초점을 맞춘 컴퓨터 통신의 여러 가지 측면에 대한 논의를 반영하는 문서를 가리킨다. 즉 RFC는 인터넷에서 사용되는 프로토콜이나 서비스에 대해 상세히 기술하거나 이에 대한 설명, 각 프로토콜의 표준화 상황 등을 정리한 문서를 지칭한다.

흔히 IETF(Internet Engineering Task Force)와 산하 운영위원회에서 정의한 인터넷 프로토콜에 대한 설명이 RFC로 발행되

고 있다. RFC에는 문서 하나 하나에 고유한 안내 번호가 부여되며 1969년 처음으로 RFC1(Host Software)이 발행된 이후 1999년 현재 RFC 2518(HTTP Ex-tensions for Dis-tributed Author-ing)까지 발행된 상태이다.

RFC는 인터넷의 표준이라기 보다는

특히의 공개와 유사하여 규격을 정하기 전에 이를 미리 일반에 공개하여 반대 의견이나 각계 의견을 모으는 절차로 이해할 수 있다.

### 2. RFC1807의 개요

과학기술의 발전과 연구 결과의 폭발적인 증가로 수많은 기술보고서가 생성되면서 연구소 나 기술보고서의 출판기관들은 우편으로 주고 받던 기술보고서의 배달과 비용문제로 어려움에 직면하게 되었고, 보고서를 수령한 후에도 정리와 검색에 많은 시간을 투자하게 되었다.

그러나 이러한 기술보고서의 출간소식을 접하기 위해 전자우편이나 파일전송의 방법을 이용하기를 원하는 기관이 늘어나면서 기술보고서를 생성하거나 다른 기관의 기술보고서를 입수하기 위해 전자메일을 사용하거나 파일전송프로토콜(FTP)을 사용하는 기관은 이들 보고서의 서지레코드의 전자적 포맷을 원하게 되었고 이를 위한 일종의 메타데이터로서 RFC1807이 탄생하게 되었다.

RFC1807은 'A Format for Bibliographic Records'로서 기술(technical)보고서를 기술(describing)하기 위한 서지레코드 포맷(형식)으로 정의할 수 있는데, 코넬대학의 Dienst 프로토콜과 스탠포드대학의 SIFT 시스템에서 제한적으로 사용되고 있다.

이 포맷의 기원은 1992년 ISI(Information Science Institute)의 Danny Cohen이 개발한 RFC1357이며 RFC1807은

#### 연재 순서

1. 메타데이터의 개요
2. DC(Dublin Core)
3. GILS(Government Information Locator Service)
4. IAFA Templates
5. MARC
6. PICS(Platform for Internet Content Selection)
7. RFC 1807 ..... 이번호
8. SOIF
9. TEI header
10. URC(Uniform Resource Characteristics)
11. Warwick Framework RDF(Resource Description Framework)
12. 메타데이터 향후 방향

RFC1357에 'handle', 'other\_access', 'keyword', 'withdraw'의 4개 필드를 추가하여 만든 개정 포맷이다.

RFC1807은 Dienst 시스템의 일부분으로 구현되었으며 코넬, 스탠포드, 캘리포니아, MIT, 카네기멜론대학 등 ARPA가 지원하는 5개의 컴퓨터과학기관이 서지레코드를 교환하기 위해 사용하고 있다. 또한 RFC1807과 USMARC형식과의 매핑을 위한 프로그램도 개발되어 있다.

RFC1807은 레코드 자체를 설명하는 알파벳 태그를 가지며 기계나 사람이 이해하기 쉽도록 매우 간단하게 설계하여 텍스트 편집기를 이용하여 쉽게 서지레코드를 생성하는 한편, 읽기 쉽고 특정 프로그램이 없이도 자유롭게 만들 수 있도록 하고 있다. 알파벳 태그는 디지털 도서관에서 온라인상의 디지털 자원에 접근하기 위한 링크로서의 역할을 하며 RFC1807은 29개 필드를 채택하여 필드의 ID를 태그로 사용하고 있다.

RFC1807은 서지레코드를 처리하는 방법이 아닌 서지레코드 포맷을 정의하고 있으며 서지 레코드가 물리적인 보고서와 온라인 디지털 도서관의 자원을 연결하는 중요한 링크의 역할을 하게 되면서 그 가치를 평가받고 있다.

### 3. 포맷

RFC1807 포맷은 필드를 지시하기 위해 스스로 각자의 요소를 설명하는 알파벳 태그를 사용한다. RFC1807이 지정하고 있는 29개 필드 가운데 'BIB-VERSION', 'ID', 'ENTRY date', 'END'의 4개 필드는 필수적(M)이며 나머지는 선택적(O)이다. 필드는 내용면에서 크게 기본기술, 주제기술, URI 관련, 관리 메타데이터, 출처, 저작권 분야로 나누어 살펴볼 수 있다.

#### ■ 기본기술

##### 1) AUTHOR (O) - 개인저자명

성부터 기입하며 편집자명도 포함한다. 개인이 아닌 위원회 등의 공동 저작인 경우에는 'CORP-AUTHOR'를 사용한다. 저자가 다수인 경우에는 이를 반복하여 표기한다.

· 표기형식 : AUTHOR:: <free-text>

· 표기예 : AUTHOR:: Finnegan, James A.

AUTHOR:: Pooh, Winnie The

##### 2) CORP-AUTHOR (O) - 위원회나 패널 등의 공동 저자

· 표기형식 : CORP-AUTHOR:: <free-text>

· 표기예 : CORP-AUTHOR:: Committee on long-range computing

##### 3) TITLE (O) - 저자가 부여한 표제

부제와 함께 완전한 표제를 기입한다.

· 표기형식 : TITLE:: <free-text>

· 표기예 : TITLE:: The Computerization of Oceanview with High Speed Fiber Optics Communication

##### 4) ORGANIZATION (O) - 발행기관명

약자가 아닌 완전한 전체 기관명을 기입한다. 이것은 ID 필드의 발행자의 통제기호와 함께 통제된다.

· 표기형식 : ORGANIZATION:: <free-text>

· 표기예 : ORGANIZATION:: Standard University, Department of Computer Science

##### 5) TYPE (O) - 발행형태

요약, 최종 프로젝트 보고서 등과 같이 발행기관이 설정한 형태를 기입한다.

· 표기형식 : TYPE:: <free-text>

· 표기예 : TYPE:: Technical Report

##### 6) CONTACT (O) - 저자의 연락처

E-mail이나 우편주소를 기입하며 공저자인 경우, 각 저자의 개별 연락처를 기입한다.

· 표기형식 : CONTACT:: <free-text>

· 표기예 : CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ., Oceanview, Kansas, 54321  
Tel: 913-456-7890 <Finnegan@cs.ouks.edu>

##### 7) DATE (O) - 발행날짜

"월 년" 혹은 "월 일, 년"의 형식으로 기입한다.

· 표기형식 : DATE:: <date>

· 표기예1 : DATE:: January 1992

· 표기예2 : DATE:: January 15, 1992

## 8) PAGES (O) - 전체 페이지수

· 표기형식 : PAGES:: &lt;number&gt;

· 표기 예 : PAGES:: 48

## 9) PERIOD (O) - 날짜범위, 기간

보고서 진행에 걸린 날짜 등에 적용된다.

· 표기형식 : PERIOD:: &lt;date&gt; to &lt;date&gt;

· 표기 예 : PERIOD:: January 1990 to March 1990

## 10) SERIES (O) - 연속간행물 표제

· 표기형식 : SERIES:: &lt;free-text&gt;

· 표기 예 : SERIES:: Communication

## 11) NOTES (O) - 주석

주석 혹은 메모로서 기술 형식에 구애받지 않는다.

· 표기형식 : NOTES:: &lt;free-text&gt;

· 표기 예 : NOTES:: This report is the full version of the paper with the same title in IEEE Trans ASSP

Dec 1976

## 12) FUNDING (O) - 재정지원기관명

· 표기형식 : FUNDING:: &lt;free-text&gt;

· 표기 예 : FUNDING:: ARPA

## 13) CONTRACT (O) - 계약 번호

· 표기형식 : CONTRACT:: &lt;free-text&gt;

· 표기 예 : CONTRACT:: MMA-90-23-456

## 14) GRANT (O) - 허가 번호

· 표기형식 : GRANT:: &lt;free-text&gt;

· 표기 예 : GRANT:: NASA-91-2345

## 15) MONITORING (O) - 감독기관명

· 표기형식 : MONITORING:: &lt;free-text&gt;

· 표기 예 : MONITORING:: ONR

## 16) LANGUAGE (O) - 보고서에서 사용한 언어

특정 언어에 대한 표기가 없으면 영어인 것으로 간주한다.

· 표기형식 : LANGUAGE:: &lt;free-text&gt;

· 표기 예 : LANGUAGE:: French

## ■ 주제기술

## 17) KEYWORD (O) - 통제 혹은 통제되지 않은 키워드

· 표기형식 : KEYWORD:: &lt;free-text&gt;

· 표기 예 : KEYWORD:: Scientific Communication  
KEYWORD:: Communication Theory

## 18) CR-CATEGORY (O) - CR(Computer Reviews)-범주

CR-범주명에는 CR-범주 색인(예, 'B.3')을 함께 기입한다. 매년 CR 1월호에는 범주의 완전 리스트와 CR 분류시스템에 대해 자세히 기재하고 있다.

· 표기형식 : CR-CATEGORY:: &lt;free-text&gt;

· 표기예1 : CR-CATEGORY:: D.1

· 표기예2 : CR-CATEGORY:: B.3 Hardware, Memory Structures

## 19) ABSTRACT (O) - 초록

길이에 대한 제한은 없지만 10,000자를 넘지 않도록 한다. 보고서가 영어로 작성되지 않았다 하더라도 초록은 영문으로 작성하는 것이 좋다. 공식적인 초록이 문서상에 없으면 서지레코드 작성자는 첫 문단의 도입 부분을 사용하게 된다.

· 표기형식 : ABSTRACT:: xxxx.....xxxx

xxxx.....xxxx

## ■ URIs

20) HANDLE (O) - 소재데이터를 찾기 위한 핸들관리시스템에서 사용하는 식별기호

핸들은 디지털 도서관내에서 저장된 디지털 자료를 확인하기 위해 사용되며 기술보고서를 전자형태로 이용할 수 있다면 핸들이 서지레코드에 나타나야 한다. 문자열은 발행자가 부여한 기술 보고서의 번호이다.

· 표기형식 : HANDLE:: hdl: &lt;naming authority&gt;/string of characters

· 표기 예 : HANDLE:: hdl:oceanview.electr/CS-TR-91-123

21) OTHER\_ACCESS (O) - URLs, URNs 등의 검색시스템에서의 여러 접근점

한 필드에 한 개의 URL이나 URN만을 제공한다.

· 표기형식 : OTHER\_ACCESS:: URL: <URL>

OTHER\_ACCESS:: URN: <URN>

· 표기예1 : OTHER\_ACCESS::URL:http://elib.

stanford.edu/Document/

STANFORD.CS:CS-TN-94-1

· 표기예2 : OTHER\_ACCESS:: URL:ftp://JUPITER.

CS.OUKS.EDU/PUBS/

computerization.txt.

## ■ 관리 메타데이터

22) BIB-VERSION (M) - 서지레코드를 작성하는데 사용되는 포맷의 버전

레코드상의 첫 번째 필드로 필수적인 요소이다. X자로 시작되는 것은 실험적인 것이며 수취인의 영구적인 데이터베이스에는 포함되지 않는다.

· 표기형식 : BIB-VERSION:: CS-TR-v2. 1

23) ID (M) - 서지레코드의 식별기호

레코드상의 두 번째 필드로 필수적인 요소이다. '/' 로 구별되는 앞부분은 발행자 ID(발행자의 통계기호)를 의미하고 뒷부분은 발행자가 부여한 발행물의 ID(예를 들면, 보고서 번호)를 가리킨다.

· 표기형식 : ID:: <publisher-ID>/(free-text)

· 표기예 : ID:: OUKS//CS-TR-91-123

24) ENTRY (M) - 서지레코드가 생성된 날짜

레코드상의 세 번째 필드로 필수적인 요소이다. ENTRY 날짜의 형식은 "월 일, 년"으로 표기한다.

· 표기형식 : ENTRY:: <date>

· 표기예 : ENTRY:: January 15, 1992

25) REVISION (O) - 서지레코드의 개정정보

개정된 날짜와 그밖의 개정과 관련된 내용을 기입한다. 발행에 대한 완전한 정보를 담고 있어야 하며 개정필드가 없으면 새 레코드가거나 개정한 적이 없는 레코드로 인식한다.

· 표기형식 : REVISION:: January 1, 1995: <free-text>

· 표기예 : REVISION:: January 1, 1995: FTP information added

26) WITHDRAW (O) - 레코드의 철회사항을 지시

문서가 더 이상 이용되지 않음을 의미하며 일부 기관에서는 레코드를 삭제하거나 또다른 기관에서는 몇 개의 필드를 제거하기도 하여 각 기관이 재량적으로 다를 수 있다. 이 필드에서는 철회된 이유를 밝혀야 한다. 다음은 서지레코드를 철회하는 예이다.

BIB-VERSION:: CS-TR-v2. 1

ID:: OUKS//CS-TR-91-123

ENTRY:: January 21, 1995

ORGANIZATION:: Oceanview University, Kansas,  
Computer Science

TITLE:: The Speed Fiber Optics Communication

REVISION:: January 21, 1995

WITHDRAW:: Withdrawn, found to be irrelevant

END:: OUKS//CS-TR-91-123

## ■ 출처/소스

27) RETRIEVAL (O) - 전문(full-text)을 얻을 수 있는 출처

출처를 포함한 서지정보의 배포는 제한할 수 있으나 이 포맷에서는 어떠한 제한사항도 마련하고 있지 않다.

· 표기형식 : RETRIEVAL:: <free-text>

· 표기예 : RETRIEVAL:: for full text with color pictures

send a self-addressed stamped envelope to

Prof. J. A. Finnegan, CS Dept,

Oceanview University, Oceanview, KS 54321

## ■ 권리의 존속기간/저작권

28) COPYRIGHT (O) - 인용된 보고서의 저작권 정보

· 표기형식 : COPYRIGHT:: <free-text>

· 표기예 : COPYRIGHT:: Copyright for the report (c) 1991, by J.A.Finnegan. All rights reserved.

Permission is granted for any academic

use of the report.

29) END (M) - 레코드가 끝났음을 표시

레코드의 마지막 요소로서 레코드 시작에 쓰였던 것과 같은 ID

를 사용한다.

· 표기형식 : END:: XXX/YYY

· 표기예 : END:: OUKS//CS-TR-91-123

이상의 각 필드의 포맷에 따른 기술사례는 <예>와 같다.

#### <예> RFC1807 기술 사례

```

BIB-VERSION:: CS-TR-v2.1
ID:: OUKS//CS-TR-91-123
ENTRY:: January 15, 1992
ORGANIZATION:: Oceanview University, Kansas, Computer Science
TYPE:: Technical Report
REVISION:: January 5, 1995: FTP access information added
TITLE:: Scientific Communication must be timely
AUTHOR:: Finnegan, James A.
CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ,
Oceanview, KS 54321 Tel: 913-456-7890
(Finnegan@cs.ouks.edu)
AUTHOR:: Pooh, Winnie The
CONTACT:: 100 Aker Wood
DATE:: December 1991
PAGES:: 48
COPYRIGHT:: Copyright for the report (c) 1991, by J. A. Finnegan.
All rights reserved. Permission is granted for any
academic use of the report.
HANDLE:: hdl:oceanview.electr/CS-TR-91-123
OTHER_ACCESS:: url:http://electr.oceanview.edu/CS-TR-91-123
OTHER_ACCESS:: url:ftp://electr.oceanview.edu/CS-TR-91-123
RETRIEVAL:: send email to Finnegan@cs.ouks.edu with fax number
KEYWORD:: Scientific Communication
CR-CATEGORY:: D. 0
CR-CATEGORY:: C. 2. 2 Computer Sys Org, Communication nets, Net
Protocols
SERIES:: Communication
FUNDING:: FAS
CONTRACT:: FAS-91-C-1234
MONITORING:: FNBO
LANGUAGE:: English
NOTES:: This report is the full version of the paper with
the same title in IEEE Trans ASSP Dec 1976
ABSTRACT::
Many alchemists in the country work on important fusion problems.
All of them cooperate and interact with each other through the scientific
literature.
This scientific communication methodology has many advantages.
Timeliness is not one of them.
END:: OUKS//CS-TR-91-123

```

## 4. 프로토콜

RFC1807 포맷과 관련된 프로토콜로는 개방형 분산 디지털 도서관 시스템을 제공하는 Dienst 프로토콜이 대표적이다.

Dienst는 디지털 도서관을 위한 개념적인 구조와 이 구조내에

서의 커뮤니케이션을 위한 프로토콜, 그리고 이 프로토콜이 실행하는 소프트웨어 시스템의 세 가지 정의로 요약할 수 있다. Dienst 시스템은 NCSTRL(Networked Computer Science Technical Report Library)의 기반으로서 개발되었지만, 프로토콜이 특정한 콜렉션(collection)으로 지정되지는 않는다.

Dienst 프로토콜이 제공하는 서비스는 다음과 같다.

① Repository Service : 디지털 자료의 저장과 이를 위한 고유 식별명을 부여한다.

② Index Service : 콜렉션의 탐색과 탐색 결과를 RFC 1807의 포맷을 사용하여 제공한다.

③ Meta Service(Contact Service) : 다른 모든 서비스의 소재 디렉토리를 제공한다.

④ User Interface Service : 이용자가 디지털 도서관에 접근할 수 있도록 해 준다.

⑤ Infor Service : 서버에 대한 일반 정보를 제공해 줄 것을 요청한다.

## 5. 핸들시스템

핸들시스템(Handle System)은 인터넷상의 디지털 객체와 다른 자원에게 핸들(handle)이라고 알려진 지속적인 식별기호를 할당, 관리, 변환하기 위한 포괄적인 시스템이다. 핸들은 URN으로 사용될 수 있고 프로토콜 집합, 이름공간(namespace), 프로토콜의 구현을 수행한다.

프로토콜은 분산된 컴퓨터 시스템이 디지털 자원의 핸들을 저장하고 이 핸들을 자원의 위치매김과 접근에 필요한 정보로 변환하도록 한다. 핸들을 바꾸지 않고도 식별된 자원의 최신 상태를 반영할 필요가 있을 경우에는 이 정보는 바뀔 수 있다. 따라서 위치와 다른 상태 정보가 변경되어도 항목의 이름은 지속될 수 있다.

핸들시스템은 많은 프로젝트에서 성공적으로 사용되고 있는데 RFC1807 개발의 배경이 된 NCSTRL(Networked Computer Science Technical Reference Library) 프로젝트를 비롯하여 미국 의회도서관(LC), 미국 국방기술정보센터(DTIC), 디지털 객체 식별기호(DOI) 프로젝트 등이 포함되어 있다.

모든 핸들은 접두어로 알려진 명명기관(Naming Authority)과 접미어로 알려진 해당 명명기관에 속하는 고유 항목 ID(Item ID)로 구성된다. 접두어, 혹은 명명기관은 행정단위를 가리키는데 이는 한번 취득되면 전세계적으로 유일하고 지속적인 것이

된다.

<그림>은 핸들시스템의 구조와 운영을 나타낸 것으로 표시된 각 과정을 설명하면 다음과 같다.

### ■ 변환(resolution)

1) 웹 브라우저와 같은 클라이언트가 인터넷이나 개별 인트라넷상에서 하이퍼링크 등을 통해 핸들과 만나게 된다. 클라이언트는 변환을 위해 핸들을 핸들시스템으로 보낸다. 이 과정은 핸들 변환 프로토콜을 본래부터 인식하고 있는 클라이언트를 통해 직접 실행되거나 그렇지 못한 클라이언트의 경우에는 프록시 서버를 통해 실행된다.

2) 핸들시스템은 여러 가지 핸들서비스의 집합으로 구성된다. 각 서비스는 수많은 서버에 물리·논리적으로 분산될 수 있으며 각 서버는 반복될 수 있다. Global Handle Registry(GHR) 서비스는 모든 공공 지역서비스(Local Service)의 위치와 이름공간의 수행 임무를 알고 있어야 하며, 각각의 지역서비스는 GHR에 접근하는 방법을 알고 있어야 한다.

3)4) 각각의 핸들은 하나 혹은 그 이상의 유형화된 데이터와 연결될 수 있다. <그림>에서는 10.123/456이란 핸들이 URL과 RAP라고 불리는 새로운 프로토콜과 연결되어 있다. 이것은 클라이언트에게 되돌려지는 정보이다. 여기에서 다중 URL과 같이 같은 데이터 형태의 다수 예를 하나의 핸들과 연결할 수 있다는 것을 염두해 둘 필요가 있다. 핸들시스템은 순수한 변환시스템이며 클라이언트의 변환 정보 처리여부를 제한하고 있지 않기 때문에 각종 응용을 위한 최대한의 유연성을 가지고 있다.

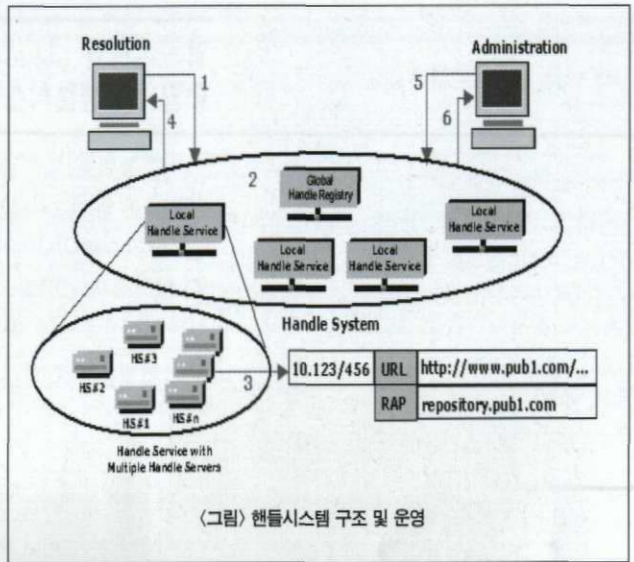
### ■ 관리(Administration)

5) 관리 클라이언트는 핸들을 생성하고 연결된 변환정보의 추가 및 수정 작업을 수행한다. 이 작업은 본래부터 핸들 관리 프로토콜을 인식하고 있는 클라이언트를 직접 사용하여 실행하거나 그렇지 못한 클라이언트의 경우 프록시 서비스를 통해 실행한다.

6) 시스템은 적절하게 성공이나 실패 메시지를 보낸다. 일괄 포맷은 다량의 추가와 수정을 위해 정의된 것이다.

## 6. 구현사례

RFC1807은 ARPA(Advanced Research Projects Agency)가 지원하는 코넬, 스탠포드, 캘리포니아, MIT, 카네기멜론대학의 5



개 컴퓨터과학기관의 서지레코드 교환을 위해 사용되고 있다. 그리고 코넬대학의 개방형 분산 디지털 도서관을 위한 Dienst 시스템의 Dienst 프로토콜과 스탠포드대학의 SIFT시스템(뉴스그룹)에서도 사용되고 있다.

또한 비영리적이고 교육적인 목적을 위해 정부연구소, 기술 단체들의 컴퓨터 부문 전문가와 사서들이 만든 온라인상의 자료 교환을 위한 프로젝트인 NCSTRL이라는 분산된 디지털 도서관 환경에서도 구현되어 NCSTRL의 메타데이터 포맷으로 RFC1807을 채택하고 있다.

RFC1807이 구현된 NCSTRL은 전세계적으로 30개 이상의 컴퓨터공학 관련 기관 및 단체가 참여하고 있으며 이용이 용이하여 연구자, 학생, 공학자, 그 밖의 관련자들이 인터넷을 통해 쉽게 이용할 수 있다.(<http://www.ncstrl.org>)

RFC1807은 분산된 컴퓨터공학 디지털 도서관을 구현하는데 필요한 서지레코드 포맷이며 디지털 도서관 시대가 도래함에 따라 그 이용 요구가 더욱 높아지리라 예상된다. ☞

### 〈참고문헌〉

- 문헌정보처리연구회 편. 1998. 메타데이터의 형식과 구조, 서울: 문헌정보처리연구회.
- "RFC1807 : A Format for Bibliographic Records". (<http://www2.ecst.csuchico.edu/~jacobsd/bib/formats/rfc1807.html>).
- "RFC1807 A Format for Bibliographic Records". June 1995. (<ftp://ftp.isi.edu/in-notes/rfc1807.txt>).
- "The Handle System". CNRI. (<http://www.handle.net/overviews/hsversion4.html>).